

2019 NCKU CSIE PhD Qualifying Exam

Biostatistics

1. Data collection [20%]

1.1 In data collection, please state the meaning of accuracy and precision.

1.2 There are two sphygmomanometers (血壓計) named S1 and S2. S1 has high accuracy and low precision. S2 has low accuracy and high precision. In a medical study, if you have to collect blood pressure data from patients, would you use S1 or S2? Please explain why. How would the imperfection in the sphygmomanometer affect collected data? Any thought to correct collected data?

2. Probability distribution [30%]

2.1 In 2010, a tree disease called Sudden Oak Death (SOD) infected various plants in California. From previous studies, it is known that the probability for a tree in Sierra Nevada (內華達山脈) to be infected by SOD is 0.1. For this study, if a research team randomly selected 5,000 trees in Sierra Nevada, please model the number of trees among 5,000 to be infected by SOD with a probability distribution. Please clearly define parameters of the probability distribution.

2.2 Often, to facilitate calculation, we may use approximation of a probability distribution. In 2.1, the total number of 5,000 inspected trees is large and it may not be easy to calculate related probability values. Please find a continuous probability distribution to approximate the distribution in 2.1. Please clearly define parameters of the continuous probability distribution and examine whether the approximation is appropriate.

2.3 Using results in 2.2, please explain how to find out the probability that between 480 to 520 trees among 5,000 trees would be infected by SOD. Please explain any correction required to calculate the probability. There is no need to find out the exact probability value (不需算出機率值，只需列出算式).

3. Confidence Interval [10%]

To find out the average body weight μ (in kg) of all male diabetes patients in Taiwan, a scientist randomly selects a sample of n male diabetes patients ($n > 30$) and measures their body weight. Please write the probability distribution of the sample mean \bar{X} (in kg) given that the standard deviation of body weight of all male diabetes patients in Taiwan is σ (in kg). From there, please derive 95% confidence interval of μ .

4. Hypothesis testing [30%]

4.1 In hypothesis testing, please explain Type-I error, Type-II error and power of the test.

4.2 It is known that the recovery period for fracture patients follows normal distribution with its standard deviation of σ (in days). In a study, a doctor collected a sample of size n with sample mean of \bar{X} (in days), please help him/her to conduct a hypothesis test of whether the true average recovery period for all fracture patients is μ_0 (in days). Please describe the general procedures for hypothesis testing. There is no need to find out exact probability values (不需算出機率值，只需詳述假設檢定之步驟).

5. Regression [10%]

Please explain the concept of linear regression. Please describe the general procedure to fit a linear regression model between an independent variable X and a dependent variable Y . How do we usually interpret the relationship between X and Y ?